Week 1: Elements of probability theory

1.1. Basic probability theory

Sample space: space of possible outcomes

Ex: Flipping a coin once: $S = \{H, T\}$
" " " twice: $S = \{HH, HT, TH, TT\}$

Probability distribution on S: $\{p_i\}_{i=1}^{|S|}$ (i) $p_i \geq 0$, $i \in S$
(ii) $\sum_{i=1}^{|S|} p_i = 1$

Event: Subset of sample space:

$$P(A) = \sum_{i \in A} p_i$$

$(P(S) = 1)$

Combination of events:

$$P(A \cup B) = \text{Prob}(A \text{ or } B)$$

$$P(A \cap B) = \text{Prob}(A \text{ and } B) = P(A, B) \quad \text{Joint probability}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Conditional probability: $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A, B)}{P(B)}$

Bayes Rule: $P(A|B) = P(B|A) \dfrac{P(A)}{P(B)}$

(Remember: $P(A,B) = P(B,A)$ and $P(A,B) = P(A|B)P(B)$)

Independent events: $P(A, B) = P(A)\,P(B)$

$$\Rightarrow P(A|B) = P(A)$$

1.2. Random variables

Random variable : Quantity X such that

i) Possible values $x_1, \dots, x_n$ (discrete case)

ii) Probabilities $p_1, \dots, p_n$

Notation: $p_i = P(X = x_i)$

$$p_i = P(\underset{\substack{\uparrow \\ \text{RV}}}{X} = \underset{\substack{\uparrow \\ \text{value}}}{x_i})$$

Bernoulli RV: $X \sim \text{Bernoulli}(p)$ $\quad P(X=1) = p$, $\quad P(X=0) = 1-p$ $\quad S = \{0,1\}$

Expectation : $E[X] = \sum_i x_i p_i = \sum_i x_i P(X = x_i)$ (discrete RV)

$= \langle X \rangle = \bar{X}$

Property: a, b constants : $E[aX+b] = a E[X] + b$

Variance : $\text{var}(X) = E[(X - E[X])^2]$ $\quad$ ~ Spreading of X

$= \langle (X - \bar{X})^2 \rangle$

$= E[X^2] - E[X]^2 = \overline{X^2} - \bar{X}^2$

Standard deviation : $\sigma(X) = \sqrt{\text{var}(X)}$

~~Propy~~

Property: $\text{var}(aX+b) = a^2 \text{var}(X)$

Independent RVs : $P(X = x_i, Y = y_i) = P(X = x_i) P(Y = y_i) \quad \forall x_i, y_i$

$\Leftrightarrow X \perp\!\!\!\perp Y$

Property: $X_i \perp\!\!\!\perp$: $\text{var}(\sum_i X_i) = \sum_i \text{var}(X_i)$

→ See material on covariance.

1.3. Continuous random variables,

$X \in \mathbb{R}$ > Probability density function: $p_X(x)$ or $f_X(x)$ for X such that

i) $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$

ii) $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

Probability of events (interval):

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$

$$P(X \in A) = \int_A f_X(x)\, dx \qquad P(X \in [x, x+dx]) = p_X(x)\, dx$$

> Cumulative distribution $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} p_X(x')\, dx'$

Expectation: $E[X] = \int_{-\infty}^{\infty} x\, p_X(x)\, dx$ $\qquad p_X(x) = \frac{d}{dx} F_X(x)$

Variance: $\mathrm{var}(X) = E[X^2] - E[X]^2$

n^{th}-moment: $E[X^n]$ $\qquad$ 1^{st} moment = mean

0^{th} " = normalization $= E[1] = 1$

Characteristic function:

$$G_X(k) = E[e^{ikX}] = \int_{-\infty}^{\infty} p_X(x)\, e^{ikx}\, dx \qquad k \in \mathbb{R}$$

$$= \text{Fourier transform of } p_X$$

$$= 1 + \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} E[X^n]$$

Generating function:

$$M_X(k) = E[e^{kX}] = \int_{-\infty}^{\infty} p_X(x)\, e^{kx}\, dx$$

$$= \text{Laplace transform of } p_X$$

$$= 1 + \sum_{n=1}^{\infty} \frac{(k)^n}{n!} E[X^n]$$

Rem: Can be used to get any moment by differentiation

Cumulant function: $C_X(k) = \ln G_X(k)$

$\Phi_X(k)$

Joint pdf: $P_{X,Y}(x,y)$ $\quad$ $P_X(x) = \int P_{XY}(x,y)\, dy$

marginals $\quad$ $P_Y(y) = \int P_{XY}(x,y)\, dx$

Independent if $P_{XY}(x,y) = P_X(x) P_Y(y)$

† Property: $E[e^{ikX} e^{ikY}] = E[e^{ikX}] E[e^{ikY}]$

1.4 Gaussian or normal RV

$X \sim \mathcal{N}(\mu, \sigma^2)$ $\quad$ $P_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$

- Mean: $E[X] = \mu$
- Variance: $var(X) = \sigma^2$
- Standard normal: $X \sim \mathcal{N}(0,1)$

~~/SX~~

Standardization: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$

Sum of Gaussians: $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ $\quad$ $X_1 \perp\!\!\!\perp X_2$

$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

$\Rightarrow X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ $\quad$ n use GF or CF

→ See exercise

Generalization: $X_i \sim \mathcal{N}(\mu, \sigma^2)$ independent

$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$

† Properties:

- $E[XY] = E[X]E[Y]$ if $X \perp\!\!\!\perp Y$
- $E[X+Y] = E[X] + E[Y]$ for $\forall X, Y$
- $E[e^{ikX} e^{ikY}] = E[e^{ikX}] E[e^{ikY}]$

$\Rightarrow G_{X+Y}(k) = G_X(k) G_Y(k)$

→ See exercises for standard pdfs + RVs

1.5 Transformation of RVs

Theorem: Let X be a continuous RV and let $Y = g(X)$, with g a differentiable function which is

i) strictly monotonically increasing (i.e. $g'(x) > 0 \quad \forall x \in \mathbb{R}$)

ii) " " decreasing ($g'(x) < 0 \quad \forall x \in \mathbb{R}$)

Then:

$$p_Y(y) = \begin{cases} p_X(g^{-1}(y)) \underbrace{\left| \frac{d}{dy} g^{-1}(y) \right|}_{\text{Jacobian}} & y \text{ such that } g^{-1}(y) \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

Proof: See notes.

Remembering the formula: $p_Y(y)\,dy = p_X(x)\,dx$

$$\Rightarrow \quad p_Y(y) = p_X(x) \frac{dx}{dy} \qquad \begin{matrix} y = g(x) \\ x = g^{-1}(y) \end{matrix}$$

General formula if not monotonic:

Example 1: (1.39 in notes)

$X \sim \mathcal{N}(\mu, \sigma^2)$

$Y = aX + b$

$Y = g(X) = aX + b$ $\quad\quad$ $g(x) = ax + b$

$X = g^{-1}(X)$ $\quad\quad$ $g^{-1}(y) = \frac{y-b}{a}$

$\frac{d}{dy} g^{-1}(y) = \frac{1}{a}$ $\quad\quad$ Jacobian

$$\Rightarrow p_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2} \right) \frac{1}{|a|}$$

$$= \frac{1}{\sqrt{2\pi a^2 \sigma^2}} \exp\left(- \frac{(y - a\mu - b)^2}{2\sigma^2 a^2} \right)$$

$$= \frac{1}{\sqrt{2\pi\sigma'^2}} e^{-(y-\mu')^2 / 2\sigma'^2}$$

$\mu' = a\mu + b$

$\sigma'^2 = a^2\sigma^2$

See Standardization.

Example 2 : (Sec. 1.5 in notes)

$X \sim N(\mu, \sigma^2)$

$Y = e^X \geq 0$

So $\log Y = X \sim N(\mu, \sigma^2)$ $\qquad$ $Y =$ lognormal

Calc: $g(x) = e^x$

$g^{-1}(y) = \log y$

$\frac{d}{dy} g^{-1}(y) = \frac{1}{y} >$ for $y > 0$ (monotonic)

$$\Rightarrow p_Y(y) = \begin{cases} p_X(g^{-1}(y)) \left| \frac{dg^{-1}}{dy} \right| & y > 0 \\ 0 & y \leq 0 \end{cases}$$

$$= \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2 y^2}} e^{-(\log y - \mu)^2 / 2\sigma^2} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

→ See exercises

1.6. Central limit theorem (CLT)

Gaussian case:

$X_1, \ldots, X_n$ ~~independent~~ iid $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$\Rightarrow S_n = \sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

$$\Rightarrow \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \sim \mathcal{N}(0,1)$$

i.e. $$\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \sim \mathcal{N}(0,1)$$

Actually much more general!

Theorem (CLT): Let $X_1, \ldots, X_n$ be iid RVs with mean $E[X_i] = \mu$ and variance $\mathrm{Var}(X_i) = \sigma^2$. Then

$$\lim_{n \to \infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \leq x \right) = \Phi(x) = P_{\text{Gaussian}}(X \leq x)$$

i.e.

$$\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \xrightarrow[\text{dist}]{n \to \infty} X \sim \mathcal{N}(0,1)$$

Proof: See notes.